

What is claimed is:

1. A method comprising:

5 obtaining performance data for software that has executed in a data processing system, wherein the performance data comprises instruction addresses and corresponding performance information;

obtaining dump information from the data processing system, wherein the dump information comprises the instructions and corresponding instruction addresses;

10 automatically identifying common code segments in the dump information, wherein a common code segment comprises an ordered set of multiple instructions that appears multiple times in the dump information; and

generating aggregate performance data for the common code segments, based at least in part on the instruction addresses associated with the common code
15 segments from the dump information, the instruction addresses from the performance data, and the corresponding performance information from the performance data.

2. A method according to claim 1, wherein:

20 the operation of obtaining performance data comprises obtaining performance data for instructions generated by a dynamic compiler; and

the operation of generating aggregate performance data for the common code segments comprises generating aggregate performance data for common code segment generated by the dynamic compiler.

25 3. A method according to claim 1, wherein the operation of identifying common code segments in the dump information comprises:

selecting a candidate code segment from the dump information;

determining whether the candidate code segment occurs multiple times in the dump information; and

30 identifying the candidate code segment as a common code segment in response to determining that the candidate code segment occurs multiple times in the dump information.

4. A method according to claim 1, wherein the operation of identifying common code segments in the dump information comprises:

selecting a candidate code segment from the dump information;

determining whether the dump information includes at least one additional

5 absolute match for the candidate code segment; and

identifying the candidate code segment as a common code segment in response to determining that the dump information includes at least one additional absolute match for the candidate code segment.

10 5. A method according to claim 1, wherein the operation of identifying common code segments in the dump information comprises:

selecting a candidate code segment from the dump information;

identifying elements in the candidate code segment as significant;

determining whether the dump information includes at least one additional

15 match for the candidate code segment, wherein the additional match comprises instructions with elements matching the significant elements in the candidate code segment; and

identifying the candidate code segment as a common code segment in response to determining that the dump information includes at least one additional match for the

20 candidate code segment.

6. A method according to claim 1, wherein the performance information comprises one or more measurements selected from the group consisting of:

execution time data for individual instructions; and

25 cache miss data for individual instructions.

7. A method according to claim 1, wherein:

the operation of identifying common code segments in the dump information comprises identifying at least first and second common code segments; and

30 the operation of generating aggregate performance data for the common code segments comprises:

collecting performance data for multiple instances of the first common code segment;

generating aggregate performance data for the first common code segment, based at least in part on the performance data for the multiple instances of the first common code segment;

collecting performance data for multiple instances of the second common code
5 segment; and

generating aggregate performance data for the second common code segment, based at least in part on the performance data for the multiple instances of the second common code segment.

10 8. A method according to claim 7, wherein the operation of generating aggregate performance data for the common code segments comprises:

collecting performance information corresponding to instruction addresses for substantially all instances of the common code segment in the dump information.

15 9. An apparatus, comprising:

a machine accessible medium; and

software encoded in the machine accessible medium, wherein the software, when executed by a processing system, performs operations comprising:

obtaining performance data for software that has executed in a data
20 processing system, wherein the performance data comprises instruction addresses and corresponding performance information;

obtaining dump information from the data processing system, wherein the dump information comprises the instructions and corresponding instruction addresses;

automatically identifying common code segments in the dump information,
25 wherein a common code segment comprises an ordered set of multiple instructions that appears multiple times in the dump information; and

generating aggregate performance data for the common code segments, based at least in part on the instruction addresses associated with the common code segments from the dump information, the instruction addresses from the performance
30 data, and the corresponding performance information from the performance data.

10. An apparatus according to claim 9, wherein:

the operation of obtaining performance data comprises obtaining performance data for instructions generated by a dynamic compiler; and

the operation of generating aggregate performance data for the common code segments comprises generating aggregate performance data for common code
5 segment generated by the dynamic compiler.

11. An apparatus according to claim 9, wherein the operation of identifying common code segments in the dump information comprises:

selecting a candidate code segment from the dump information;
10 determining whether the candidate code segment occurs multiple times in the dump information; and
identifying the candidate code segment as a common code segment in response to determining that the candidate code segment occurs multiple times in the dump information.

15

12. An apparatus according to claim 9, wherein the operation of identifying common code segments in the dump information comprises:

selecting a candidate code segment from the dump information;
determining whether the dump information includes at least one additional
20 absolute match for the candidate code segment; and
identifying the candidate code segment as a common code segment in response to determining that the dump information includes at least one additional absolute match for the candidate code segment.

25 13. An apparatus according to claim 9, wherein the operation of identifying common code segments in the dump information comprises:

selecting a candidate code segment from the dump information;
identifying elements in the candidate code segment as significant;
determining whether the dump information includes at least one additional
30 match for the candidate code segment, wherein the additional match comprises instructions with elements matching the significant elements in the candidate code segment; and

identifying the candidate code segment as a common code segment in response to determining that the dump information includes at least one additional match for the candidate code segment.

- 5 14. An apparatus according to claim 9, wherein the performance information comprises one or more measurements selected from the group consisting of:
execution time data for individual instructions; and
cache miss data for individual instructions.

- 10 15. An apparatus according to claim 9, wherein:
the operation of identifying common code segments in the dump information comprises identifying at least first and second common code segments; and
the operation of generating aggregate performance data for the common code segments comprises:
15 collecting performance data for multiple instances of the first common code segment;
generating aggregate performance data for the first common code segment, based at least in part on the performance data for the multiple instances of the first common code segment;
20 collecting performance data for multiple instances of the second common code segment;
generating aggregate performance data for the second common code segment, based at least in part on the performance data for the multiple instances of the second common code segment.

- 25 16. An apparatus according to claim 15, wherein the operation of generating aggregate performance data for the common code segments comprises:
collecting performance information corresponding to instruction addresses for substantially all instances of the common code segment in the dump information.

- 30 17. A system, comprising:
a processor;
a machine accessible medium responsive to the processor; and

instructions in the machine accessible medium, wherein the instructions, when executed by the processor, perform operations comprising:

obtaining performance data for software that has executed in a data processing system, wherein the performance data comprises instruction addresses
5 and corresponding performance information;

obtaining dump information from the data processing system, wherein the dump information comprises the instructions and corresponding instruction addresses;

automatically identifying common code segments in the dump information, wherein a common code segment comprises an ordered set of multiple instructions
10 that appears multiple times in the dump information; and

generating aggregate performance data for the common code segments, based at least in part on the instruction addresses associated with the common code segments from the dump information, the instruction addresses from the performance data, and the corresponding performance information from the performance data.

15

18. A system according to claim 17, wherein:

the operation of obtaining performance data comprises obtaining performance data for instructions generated by a dynamic compiler; and

the operation of generating aggregate performance data for the common code
20 segments comprises generating aggregate performance data for common code segment generated by the dynamic compiler.

19. A system according to claim 17, wherein the operation of identifying common code segments in the dump information comprises:

25 selecting a candidate code segment from the dump information;

determining whether the candidate code segment occurs multiple times in the dump information; and

identifying the candidate code segment as a common code segment in response to determining that the candidate code segment occurs multiple times in the
30 dump information.

20. A system according to claim 17, wherein:

the operation of identifying common code segments in the dump information comprises identifying at least first and second common code segments; and

the operation of generating aggregate performance data for the common code segments comprises:

5 collecting performance data for multiple instances of the first common code segment;

 generating aggregate performance data for the first common code segment, based at least in part on the performance data for the multiple instances of the first common code segment;

10 collecting performance data for multiple instances of the second common code segment; and

 generating aggregate performance data for the second common code segment, based at least in part on the performance data for the multiple instances of the second common code segment.